Validation of Genomic Predictions in German Holsteins for All Traits

S. Rensing and Z. Liu

Vereinigte Informationssysteme Tierhaltung w.V. (vit), Heideweg 1, 27283 Verden (Aller), Germany

Abstract

Genomic enhanced breeding values (**gEBV**) for Holsteins are official in Germany since August 2010. Independent gEBV in April 2010 of 1,374 German A.I. bulls derived from the validation for the ICAR/Interbull test for the German Holstein genomic evaluation system were compared with their official daughter based gEBV from August 2011. Traits were the milk production index **RZM**, total conformation index **RZE**, relative breeding value somatic cell score **RZS**, combined longevity **RZN** and daughter fertility index **RZR**. For all traits differences between means for gEBV without and with daughter information were very small indicating no bias. The same was observed for the top-100 bulls for RZM whereas the top bulls for the other traits seemed to be slightly overestimated by 0.1 (RZE) to 0.2 (RZN) genetic standard deviations (s_g). Standard deviation of difference (**SD**) ranged from below 0.5 for RZS and RZR to 0.68 s_g for RZM. There were indications that gEBV without daughters predict later highly reliable daughter based gEBV better than the first daughter based proofs.

Keywords: Holsteins, genomic predictions, validation

Introduction

Genomic enhanced breeding values are official in Germany since August 2010. Legal basis is the official recognition of the German genomic evaluation system for Holsteins by Interbull carried out for protein kg. On a test basis the German genomic evaluation system for Holsteins met the developed criteria for the other milk production traits (milk kg, fat kg) and the selected traits from the non-production trait complexes (SCS, longevity, stature, non-return-rate cows) in June 2011, too. For the Interbull validation the youngest daughter proven bulls were excluded from the training set for genomics and their gEBV were calculated without daughter information. These independent gEBV (equivalent to those for young candidates) in Interbull validation were compared with deregressed proofs from the early classical genetic evaluation. For breeders and breeding organizations the comparison to official gEBV including (many) daughter and genomic information may be more relevant. Beside averages for the entire group of validation bulls individual differences and results for top bulls are important.

Material and Methods

Genomic enhanced breeding values without daughter information (candidate gEBV) were derived from the Interbull validation for milk production traits carried out on the data basis April 2010. Through the exchange of SNP data between the EuroGenomics partners from France, Germany, Scandinavia and The Netherlands the training set included that time 15,871 Holstein A.I. bulls with MACE proofs for protein kg and minimum 10 EDC (effective daughter contribution) on German scale. The youngest 1,374 bulls with proofs in Germany (born 2003/2004) were defined as validation bulls. Their candidate gEBV for all traits were calculated based on SNP effects derived from a reduced training set with the 14,497 remaining Holstein bulls. Applied methods for deriving direct genomic values are described by Liu et al. 2011 and in the GENO form for Germany at Interbull website (www.interbull.org). Assumed polygenic variance varies from 1% to 20% for individual traits. All gEBV were calculated as applied in routine evaluation since December 2010. Classical information in candidate gEBV is the sire pedigree index.

Candidate gEBV were compared with the official published daughter based gEBV from August 2011. Because of the timespan since April 2010 validation bulls had at least their 5th daughter based proof for the milk production traits and meanwhile daughter information included for all traits even for the later traits like direct longevity. Yearly base shift for all non-production traits and indices was corrected for. The difference between the two gEBV is calculated as Aug.-2011 minus April-2010 proof so that positive values indicate an increase of gEBVs after inclusion of daughter information. Results are presented for the main indices. Except for somatic cell count these are indices including several traits. Details can be found in the description of breeding values on www.vit.de. The scale was converted to average=0 and s=1.

Results

Figure 1 shows the results for the milk production index RZM (70% protein kg, 25% fat kg, 5% protein %). The daughter based gRZM in August 2011 include in average information from 122.3 daughters resulting in 94.2% average reliability. In the figures dots above the diagonal represent bulls that increased gRZM after inclusion of daughter information. On average there is no difference between the former candidate gRZM with 73% reliability and current highly reliable proofs including many daughters ($\pm$ 0.000 s_g) and EBV correlation is 0.71. SD of the differences between candidate gRZM and daughter gRZM is 0.677 s_g.

Realized reliability of candidate somatic cell score gRZS is the highest of all traits (77%) and correlation to the daughter based gRZS (122.3 daugh.) with 89.6% reliability is 0.80. The very good prediction results in only 0.480 s_g SD of the difference between the two gEBV (figure 2). With +0.030 s_g there is no bias in the average of the candidate gEBV.

Average reliability for the gRZE (total conformation index including 18 linear traits and 4 scores) based on 65.9 daughters is with 81.6% slightly lower compared to gRZM and gRZS. Nevertheless increase of reliability is higher (candidate gRZE 56% rel.). Average

difference is +0.019 s_g, SD of difference 0.595 s_g and correlation between gEBV 0.64.

Reliability of daughter based gEBV is lower for longevity and daughter fertility because heritability is lower and information is available later. The average combined longevity gRZN in August 2011 includes information from 92.5 daughters for direct longevity resulting in 65.1% reliability and a correlation with the candidate gEBV of 0.64. The SD of the difference is 0.584 s_g (figure 3) and the average difference shows no substantial bias (+0.028 s_g).

Daughter fertility index RZR has still limited reliability (53.0%) even though information from 94.1 daughters for Non-Return-Rate-56 is included. Therefore there is an autocorrelation because the daughter based gRZR includes still significant genomic information and SD of the difference (0.463 s_g) is relatively low. With +0.058 s_g daughter based gRZR is slightly higher than candidate gRZR.

For breeding decisions the prediction of the highest bulls selected on candidate gEBV is more important than average predictability.

Table 1. Comparison of candidate gEBV April-2010 with daughter based gEBV Aug.-2011 for all and top-100 bulls (in s_g).

Trait	diff. all	(SD)	top-100	(SD)
gRZM	0.000	(0.677)	-0.041	(0.689)
gRZS	0.019	(0.595)	-0.105	(0.613)
gRZE	0.030	(0.480)	-0.125	(0.425)
gRZN	0.028	(0.584)	-0.233	(0.482)
gRZR	0.058	(0.463)	-0.143	(0.425)

In table 1 results for all 1,374 validation bulls are compared with the top-100 selected on candidate gEBV for each trait (best 7.3%). For production index gRZM the results for the top-100 are very similar. For the non-production traits the top-100 were slightly overestimated in the range of 0.1 to 0.2 s_g.

The above described validation compares candidate gEBV with daughter gEBV after several daughter evaluations with already relative high classical reliability. In practical breeding farmers and especially breeding organizations mostly will make their selection

among daughter proven bulls earlier. Therefore the validation was repeated with the official gEBV from the previous publications in April-2011 and December-2010 (table 2). Comparison to first published gEBV in Aug.-2010 was not included because the methodology of calculating gEBV was different that time.

Table 2. Validation over time (daughter gEBV - candidate gEBV), all bulls (in s_g).

Trait	12-2010	04-2011	08-2011
gRZM	0.043	0.030	0.000
gRZS	0.030	0.023	0.019
gRZE	0.055	0.045	0.030
gRZN	0.170	0.171	0.028
gRZR	0.088	0.068	0.058

Table 2 indicates that average differences became smaller over time. The candidate gEBV on average predicted the later daughter proofs better than early daughter proofs. Most obvious is the change for longevity (RZN) where average difference only in 08-2011 became close to zero. In 08-2011 most of the validation bulls born in 2004 have many daughters in 2nd lactation and therefore information on many daughters even for longevity.

Discussion

This validation study focused on the prediction of later gEBV based on many daughters even in 2nd lactation. Choosing gEBV as target implies that genomic information is still included. But the impact on gEBV is limited because reliability of classical EBV is higher than direct genomic value (and absolutely high for at least production, SCS and conformation). Furthermore the validation bulls are meanwhile part of the training set and current direct genomic values are no longer independent and forced towards the classical EBV. On the other hand gEBV are the only published values and farmers and breeding organizations have to decide on these.

German Holstein genomics prove to be very realistic in average for all trait complexes. For the top bulls per trait who are in the focus of selection this is true, too, even though there is a slight tendency of overestimation of candidate gEBV, i.e. regression of candidate gEBV on daughter EBV is <1. Separate validation on top bulls is important to give farmers and breeding organizations trust in genomics on national level and to ensure that there will be no bias in across country comparison for top bulls.

Presented validation results over time indicate that candidate gEBV predict the average of highly reliable later daughter based gEBV better compared to early daughter gEBV. One reason could be bias in the early daughter proofs from animal models that are still significantly influenced by the parent average. At least for production traits parent average tend to be overestimated due to special treatment of bull dams (Rensing *et al.,* 2010). Because farmers and breeding organizations tend to validate the quality of genomics on early daughter based gEBV (for production, conformation, SCS) for the highest bulls this may cause acceptance problems for genomics due to biased early daughter based proofs.

Literature

Liu, Z., Seefried, F., Reinhardt, F., Rensing, S., Thaller, G. & Reents, R. 2011. Impacts of both reference population size and inclusion of a residual polygenic effect on the accuracy of genomic prediction. *Genetics Selelection Evoution, 43,* 19.

Rensing, S., Pasman, E. & Reinhardt, F. 2010. Best use of conventional EBV of bull dams and combination with direct genomic values. *Interbull Bulletin 40,* 123-126.

Figure 1. Comparison of daughter based gEBV for milk production index RZM in August 2011 (∅ 122.3 daughters, 94.2% reliability) with candidate gEBV without daughter information from April 2010 for 1,374 validation bulls (scale s_g).
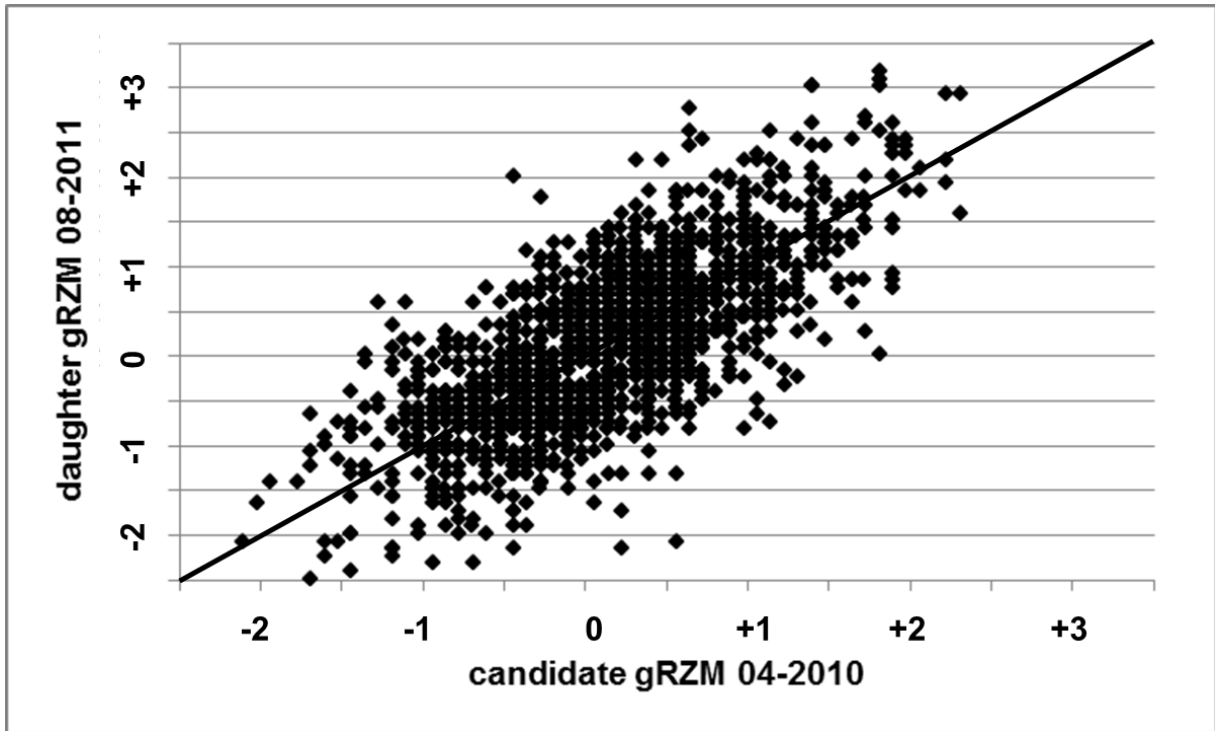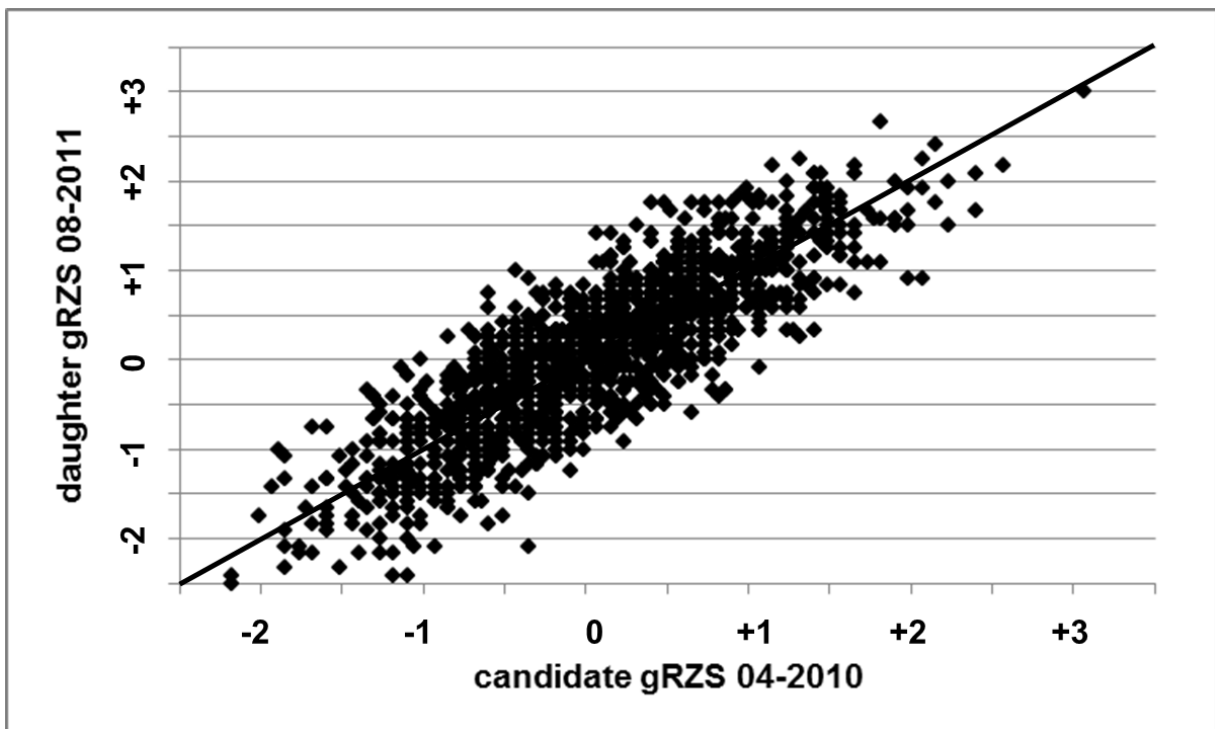

Figure 2. Comparison of daughter based gEBV for relative breeding value somatic cell score RZS (sign reversed scale: positive values = low cell count) in August 2011 (∅ 122.3 daughters, 89.6% reliability) with candidate gEBV without daughter information from April 2010 for 1,374 validation bulls (scale s_g).

Figure 3. Comparison of daughter based gEBV for total conformation index RZE in August 2011 (∅ 65.9 daughters, 81.6% reliability) with candidate gEBV without daughter information from April 2010 for 1,374 validation bulls (scale s_g).
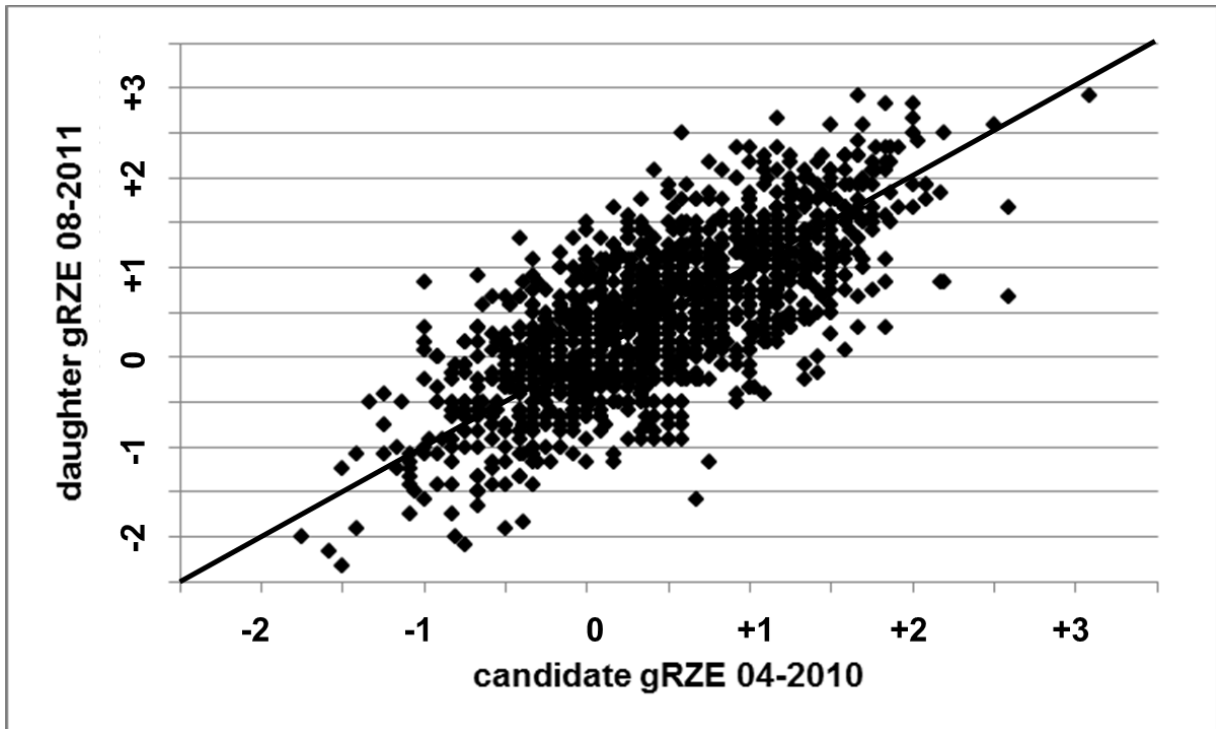

Figure 4. Comparison of daughter based gEBV for combined longevity RZN in August 2011 (∅ 92.5 daughters, 65.1% reliability) with candidate gEBV without daughter information from April 2010 for 1,374 validation bulls (scale s_g).
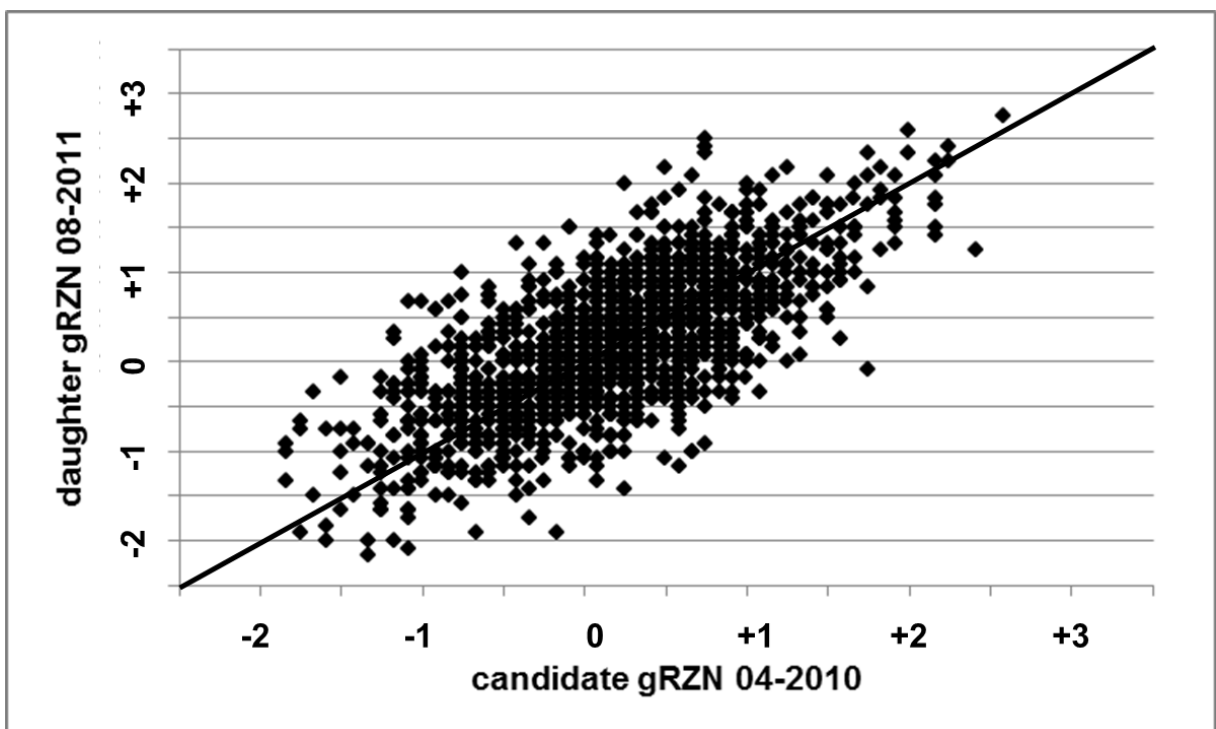
Figure 5. Comparison of daughter based gEBV for daughter fertility index RZR in August 2011 ($\varnothing$ 94.1 daughters NRR cows, 53.0% reliability RZR) with candidate gEBV without daughter information from April 2010 for 1,374 validation bulls (scale s_g).